Published in final edited form as:

Nat Biotechnol. 2007 August ; 25(8): 939–943. doi:10.1038/nbt1321.

Molecular breeding of polymerases for amplification of ancient DNA

Marc d'Abbadie¹, Michael Hofreiter², Alexandra Vaisman³, David Loakes¹, Didier Gasparutto⁴, Jean Cadet⁴, Roger Woodgate³, Svante Pääbo², and Philipp Holliger¹

¹MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

²Max Planck Institute for Molecular Anthropology, Department of Evolutionary Genetics, Deutscher Platz 6, D-04103 Leipzig, Germany

³Section on DNA Replication, Repair and Mutagenesis, Building 6, Room 1A13, National Institutes of Child Health and Human Development (NICHD), National Institutes of Health (NIH), Bethesda, MD 20892-2725, USA

⁴Laboratoire "Lésions des Acides Nucléiques", SCIB/UMR-E 3 (CEA/UJF), Département de Recherche Fondamentale sur la Matière Condensée, Commissariat à l'Énergie Atomique (CEA)/Grenoble, F-38054 Grenoble Cedex 9, France

In the absence of repair, lesions accumulate in DNA. Thus, DNA persisting in specimens of paleontological, archaeological, or forensic interest is inevitably damaged¹. We describe a strategy for the recovery of genetic information from damaged DNA. By molecular breeding² of polymerase genes from the genus *Thermus* (*Taq* (*T. aquaticus*), *Tth* (*T. thermophilus*), *Tfi* (*T. flavus*)) and CSR3, 4 selection we have evolved polymerases that can extend single, double and even quadruple mismatches, process non-canonical primer-template duplexes and bypass lesions found in ancient DNA such as hydantoins and abasic sites. Applied to the PCR amplification of 47,000–60,000 years old cave bear DNA, they outperformed *Taq* DNA polymerase by up to 150% and yielded amplification products at sample dilutions, where *Taq* no longer yielded any. Our results demonstrate that engineered polymerases can expand the recovery of genetic information from Pleistocene specimens and may benefit genetic analysis in paleontology, archaeology and forensic medicine.

Ancient DNA sequences have been isolated from a wide variety of sources¹ and have provided information about human migration⁵, animal and crop domestication^{6, 7} and the genetic relationship between modern *Homo sapiens* and its closest extinct relative *H. neandertalensis*^{8, 9}. However, even under optimal burial conditions, DNA damage is progressive and either limits the length of continuous sequence that can be recovered or renders even well-preserved specimens unproductive despite the demonstrable presence of DNA (by hybridization)¹⁰. We reasoned that genetic information encoded in such samples may not be lost but simply inaccessible due to the fact that the DNA polymerases commonly used for PCR stall at sites of damage¹¹. Polymerases capable of PCR amplification of damaged DNA would therefore have the potential to enhance the retrieval of ancient DNA sequences and, in combination with direct sequencing approaches^{8, 12}, provide a significant expansion of palaeogenomic data.

Correspondence should be addressed to PH: ph1@mrcmb.cam.ac.uk.

Author contributions: MdA contributed to library construction, CSR selection and together with MH and SP to ancient DNA amplification. MdA, PH, AV and RW contributed to polymerase characterization using template lesions synthesized by DL, DG and JC. PH contributed to the planning and design of the project and manuscript writing.

Engineering polymerases, which combine the processivity and selectivity required for PCR with a high tolerance for template damage is challenging. Furthermore, damage tolerance should be generic as detailed information about the forms of DNA damage in ancient samples is lacking (and damage may vary depending on burial conditions). Many lesions (with the exception of miscoding lesions such as uracil) abrogate base-pairing and yield distorted, non-cognate 3' structures, similar to transversion mismatches. In the case of A-family polymerases, such mismatches cause significant stalling, not just at the primer 3' end¹³, but up to four bases upstream¹⁴. In order to maximize tolerance to such distorted primer-template structures, we decided to select for polymerases capable of extending a primer 3' terminus preceded by up to four mismatched bases.

Previous library designs for polymerase evolution were based on random mutagenesis of the polymerase gene^{3, 4} or defined regions thereof¹⁵⁻¹⁷ but these proved unproductive for the selection of damage tolerant polymerases (not shown). We therefore prepared new libraries using molecular breeding². We recombined three A-family polymerase (DNA pol I) genes from the genus *Thermus*: *Taq* (*Thermus aquaticus*), *Tth* (*T. thermophilus*) and *Tfl* (*T. flavus*) to create library 3T for selection by compartmentalized self-replication (CSR)³ (Fig. 1a and supplementary Fig. 1 online). To test library performance we first selected for CC•CC double mismatch extension.

Single C•C mismatches are extended >10⁶-fold less efficiently than matched termini by *Taq*¹³. Double mismatch extension had previously only been reported for the Y-family polymerases pol η and pol ι ^{18, 19}. Nevertheless, a single round of CSR selection of 3T produced several clones with efficient double mismatch extension. In particular, H10 (a *Tth* / *Taq* chimera with nine additional point mutations (F74I, F280L, P300S, T387A, A441V, A519V, Q536R, R679G, F699L (*Tth* numbering))) could extend primers with both double (CC•CC) as well as single (C•C; G•A) mismatches in PCR (not shown). However, H10 displayed no clear improvement in damage tolerance compared to the previously described single mismatch extension polymerase M14.

We therefore proceeded to select for extension of the more challenging quadruple mismatches GGTG•GGTG, AGGG•AGGG (primer (5'-3') • template (3'-5')) (Fig. 1a). After three rounds of CSR, we recovered a diverse set of polymerase chimeras with novel properties (see supplementary Table 1 online) including the generic ability to utilize single, double and quadruple mismatches (e.g. 3D1, Fig. 1b) or bypass template lesions such as abasic sites or hydantoins (see supplementary Table 1).

Despite their diverse properties, all the selected polymerases share a similar arrangement of gene segments. While diverging in detail, in all polymerases the N-terminal region (comprising part or all of the 5'-3' exonuclease domain) and the C-terminus derive from *Tth*, whereas the protein core derives mainly from *Taq*. Four point mutations (L33P, E76K, D145G and E822K), diverging from either *Taq*, *Tth* or *Tfl*, were present in several polymerases. One variant (3B10) had acquired a 16 amino acids C-terminal extension through a frame shift-mutation (see supplementary Figure 2 online).

We chose two polymerases 3D1 (a *Tth* / *Taq* chimera with 6 additional mutations (L33P, E76K, D145G, P552S, E775G, M777T)) and 3A10 (a *Tth* / *Taq* chimera with 8 additional mutations (E76K, E91Q, D145G, R336Q, A448T, I616M, V739M, E744G)), for detailed investigation. We first examined extension of three different quadruple mismatches: GGTG•GGTG, AGGG•AGGG (used for selection), and the unrelated TTTT•TTTT. While *Taq* was unable to extend any of the mismatches, both 3A10 and 3D1 could, but the majority of the reaction products were shorter than expected (Fig. 2a,b). Extension of matched termini generated by strand misalignment appeared to compete with direct

extension of quadruple mismatches. To assess whether extension of quadruple mismatches was possible, we performed extension of the TTTT•TTTT quadruple mismatch with a single nucleotide (dTTP) complementary to the first two template bases (dA-dA) with 3A10. Misalignment was suppressed and the TTTT•TTTT quadruple mismatch was efficiently extended by correct incorporation of two dTs (Fig. 2c). Furthermore, addition of dNTPs resulted in the further extension of the quadruple mismatch and the formation of a significant amount of full-length product.

Both 3A10 and 3D1 were capable of efficient bypass of DNA lesions relevant to ancient DNA such as abasic sites (AP), 5-hydroxy- (5Hyd) and 5-methyl-5-hydroxy-hydantoin (5MeHyd) associated with PCR failure from ancient samples²⁰, while the same lesions blocked extension by the parent polymerases *Taq* (Fig. 3a, and supplementary Fig. 3a) *Tth* or *Tfi* (not shown). Both polymerases, in particular 3A10, were also proficient at bypassing abasic sites in PCR (see supplementary Fig. 3b online). 3A10 displayed translesion synthesis of both AP as well as 5Hyd lesions up 50-fold superior to *Taq* as judged by polymerase ELISA¹⁷ (see supplementary Fig. 3c online). Both 3A10 and 3D1 displayed a similar spectrum of dNTP insertion (A>G>>T>>>C) opposite AP or 5Hyd / 5MeHyd lesions. For AP sites, deriving from depurination, bypass is either silent or leads to G → A transitions, whereas for the hydantoins, which originate from pyrimidines, bypass yields C, T → A transversion mutations. We estimated fidelity on an undamaged template to be decreased by 2-4 fold (3D1) and 7-fold (3A10) compared to *Taq* (see supplementary Fig. 4a, b online).

The ability of selected polymerases to efficiently bypass template lesions in PCR encouraged us to test their activity for the recovery of ancient DNA. We performed subsequent experiments using a blend of *Taq* with the most promising selected polymerases (3A10, 3B5, 3B6, 3B8, 3B10, 3C12, 3D1) (rather than testing individual combinations) in order to minimize wastage of precious ancient samples and maximize the chances of success. We first performed 56 PCR amplifications at limiting dilutions of ancient DNA (aDNA) derived from a 47,000 year-old cave bear (*Ursus spelaeus*) bone and scored successful amplifications for blend and *Taq* alone. We found that the blend yielded amplification products at between 2 - 5-fold lower concentrations of aDNA than *Taq* and indeed did yield amplification products at DNA concentrations, where *Taq* no longer generated any (Fig. 3b).

Normalizing PCR activity on a dilution series of “modern” DNA showed that this was not due to higher PCR efficiency of the blend. On the contrary, *Taq* appeared to be more than an order of magnitude more active at low template concentrations (of “modern” DNA) (Fig. 3b), suggesting that the blend requires more template than *Taq* to produce an equivalent PCR signal. This suggests that the measured activity of the blend for the amplification of ancient DNA is likely to represent an underestimate of its true potential. Moreover, it implies that the blend can tap into a pool of DNA molecules that are inaccessible to *Taq*, presumably because they are damaged.

To stringently exclude sample heterogeneity and stochastic variation as the source of the above effect, we performed a further 608 independent PCR amplifications from two different samples of cave bear bone (~47,000 and ~60,000 years-old respectively), and scored the number of PCR amplicons at limiting dilution (see supplementary Fig. 5 online). The blend yielded a larger number of amplicons (8-150%) than *Taq* in all but one experiment (Table 1), confirming previous results.

We cloned amplicons from experiment five (Table 1), sequenced independent clones (blend (28), *Taq* (26)), and analyzed their sequences for systematic and sporadic errors. As ancient DNA PCR products may arise from a single template molecule, systematic errors (i.e.

deviations for the consensus sequence occurring in all clones from one PCR) mostly arise from lesions in the ancient DNA template. Sporadic errors, in contrast, occur at any point during the PCR amplification. Systematic errors therefore reflect lesion bypass, while sporadic errors largely reflect polymerase fidelity in PCR.

There appeared to be a higher incidence of systematic errors in the blend PCRs (5 blend / 3 *Taq*). Although numbers are small, we speculate that the difference may arise from the amplification products deriving from bypass of previously blocking lesions, broadly consistent with ca. 30% higher amplification success in experiment 5 (Table 1).

Blend amplicons also showed an 3.25-fold increase in sporadic errors (14 blend / 4 *Taq*). In order to better determine polymerase fidelity we sequenced further amplicons (300bp; 10 blend / 15 *Taq*) deriving from a different region of the cave bear mitochondrial genome. We found a 3.75-fold increase in sporadic errors (15 blend / 6 *Taq*), presumably reflecting the reduced fidelity of blend polymerases such as 3D1 and 3A10 (see supplementary Fig. 4a,b online).

The here described polymerases are unique in combining multiple lesion bypass (e.g. two abasic sites in PCR (see supplementary Fig. 2a online) with robust PCR activity. Together with their ability to process non-cognate primer duplexes, this may contribute to their ability to enhance the recovery of ancient DNA sequences

Further improvements in polymerase performance appear possible but may depend on increased insights into the bottlenecks of ancient DNA recovery. We evolved bypass to just two classes of lesions, which are known to occur in ancient DNA. Abasic sites are generated by spontaneous depurination or depyrimidination and as the end product of a number of oxidation-induced DNA damage pathways^{21, 22}. High levels of oxidized pyrimidines such as 5-hydroxy-5-methylhydantoin and 5-hydroxyhydantoin have been found in ancient samples and associated with PCR failure²⁰. However, it is possible or even likely that other poorly understood forms of damage cause PCR failure. These may include intra-strand crosslinks, which appear to be prevalent in older samples²³ and may be poorly bypassed, if at all, even by the selected polymerases with their tolerance for non-canonical primer-template duplex structures. Furthermore, PCR recovery may be limited by the presence of potent inhibitors, such as heme or polyphenolic acids produced by the decomposition of organic matter. CSR should allow the selection of polymerases with additional improvements in their ability to bypass lesions and non-canonical DNA structures. These may be combined with polymerases selected for resistance to environmental inhibitors, as demonstrated previously for heparin³.

The molecular determinants of the remarkable abilities of 3A10 or 3D1 to extend double and quadruple mismatches and process misaligned primer template structures are difficult to infer from sequence data alone. The features that are most consistently shared among the selected polymerases (for example, mutations L33P, E76K, D145G ; see supplementary Fig. 1 and 2 online) all implicate the 5'-3' exonuclease domain. In most selected polymerases, this domain derives from *Tth*. While we found that the *Tth* exonuclease domain is more thermostable than its *Taq* counterpart (FJ Ghadessy, PH, unpublished results) and may therefore promote evolvability through increased tolerance of destabilizing mutations as described²⁴, there may be other factors contributing to its universal selection. At least in the case of the previously described polymerase M1 4, we found that mutations in the 5'-3' exonuclease domain contributed substantially to mismatch extension (Md'A & PH, unpublished results), suggesting that it may contribute, in a manner yet to be understood, to the processing of non-cognate 3' ends.

In conclusion, molecular breeding and directed evolution by CSR have allowed the isolation of polymerases, which enhance the recovery of genetic material from Pleistocene specimens, presumably due to their ability to amplify damaged DNA. Polymerases such as these should improve the recovery of ancient DNA and reduce bias towards modern DNA contamination. They should also be suited for direct sequencing approaches^{8, 12} as they are pre-adapted to emulsion PCR³. Polymerases capable of amplifying damaged DNA have applications and impact beyond palaeobiology, for example in archaeology, historic and forensic medicine and the genetic analysis of clinical specimens damaged by preservatives, cancer drugs or ionizing radiation.

Materials & Methods

DNA manipulation and protein expression

Tth and *Tfl* polymerase genes were cloned from *Thermus thermophilus* (*Tth*) and *Thermus flavus* (*Tfl*) genomic DNA (DSMZ) using gene-specific primers 1-4 (oligonucleotide sequences are provided in Supplementary Materials online), cloned into pASK75 and assayed for PCR activity as described³. Polymerases libraries were prepared by molecular breeding of polymerase genes. In molecular breeding homologous genes from different organisms (orthologues) are recombined to yield a library of chimeras comprising segments of the different orthologues. Molecular breeding samples only functional diversity and therefore molecular breeding libraries often comprise a larger number of active clones than random mutant libraries. Genes for *Taq*, *Taq* mutant T83, *Tth* and *Tfl* genes were recombined using the staggered extension protocol (StEP) as described²⁵. In StEP, genes to be recombined PCR are amplified with common flanking primers but extension times that are too short to allow complete primer extension during each cycle of PCR. This promotes template-switching between homologous regions leading to effective recombination between the genes. Tuning the extension time allows some control over the length of gene segments that are swapped. Here, equal concentrations of polymerase genes were cycled 40 × (94°C 30sec, 55°C 1sec) using primers 5,6. The product was gel-purified and reamplified with primers 7,8 and cloned XbaI / SalI into pASK75 to create library 3T (1 × 10⁹ cfu, 70% active clones). Expression of polymerases for characterization and ancient DNA PCR was as described^{3, 4} but using a 16/10 Hi-Prep Heparin FF Column (Amersham Pharmacia Biotech) column to purify heat-cleared (50°C, 30min) Bugbuster (Novagen) lysate with Complete™ EDTA-free protease inhibitor cocktail (Roche). Polymerase fractions eluted around 0.3M NaCl and were concentrated and dia-filtrated into 50Mm Tris pH 7.4, 1mM DTT, 50% glycerol and stored at -20°C.

Selection and Screening

Emulsification and CSR selection were performed as described^{3, 26} using either matched primers 5, 6 or single-mismatch primers 9, 10⁴, double-mismatch primers 11, 12 and quadruple mismatch primer 13, 14 cycled 20x (94°C 1min, 50°C 1min, 72°C 8 min), reamplified with out-nested primers 5,6 or with gene specific primers 1-4,15 or combinations thereof and recloned as above. After selection rounds one and two, clones were screened by mismatch PCR (94°C 30 sec, 55°C 30 sec, 72°C 1 min) with primers 5,6 (20 cycles (20x)) or 9,10 (30x) or 11,12 (30x) or 13,14 (50x) abasic site bypass-PCR with primers 16, 17 (25x) and by polymerase ELISA as described¹⁷ but using hairpins 18, 19. Promising clones from rounds one and two were StEP shuffled and backcrossed with parent polymerase genes. Clones analyzed in more detail in this report derive from selection round three. Mutation rates were determined using the mutS ELISA assay²⁷ (Genecheck, Ft. Collins, CO) according to manufacturers instructions.

Primer extension

Synthesis of 5-hydroxy-hydantoin phosphoramidite is described in Supplementary Materials online. DNA primers and templates substrates were prepared as described²⁸ by annealing single-stranded circular DNA with ³²P-labeled primers at a 1.5:1 molar ratio in annealing buffer (50 mM Tris-HCl (pH 8), 5 mM MgCl₂, 50 µg/ml BSA, 1.42 mM 2-mercaptoethanol) for 10min at 100°C followed by slow cooling to room temperature over 2h. Annealing efficiencies were >95%, as evidenced by the different mobility of the ³²P-labeled primers before and after hybridization to the template on non-denaturing polyacrylamide gels. Primer 18 was used to study abasic and 5-hydroxy-hydantoin bypass on templates 21-23. To study bypass of 5-methyl-5hydroxy-hydantoin, primer 24 and templates 25, 26 were used. To study quadruple mismatch extension primers 27-30 and templates 31-33 were used. Standard extension reactions contained 10 nM DNA templates (expressed as primer termini); 100 µM of either all four dNTPs or each dNTP individually; 40 mM Tris-HCl at pH 8.0; 5 mM MgCl₂; 10 mM dithiothreitol, 250 µg/ml bovine serum albumin; 2.5% glycerol, and various amounts of DNA polymerases. Reactions were incubated for 10min at 65 °C unless specified otherwise. The reactions were terminated by mixing with one volume of formamide loading dye solution containing 50 mM EDTA, 0.1% xylene cyanol, and 0.1% bromophenol blue in 90% formamide. Before loading onto the gel, the reactions were denatured by heating at 100 °C for 10 min and immediately transferred into ice for 2 min. Products were resolved by denaturing polyacrylamide gel electrophoresis (8 M urea, 15% acrylamide, 3h at 2000 V) and then visualized and quantified using a Fuji image analyzer FLA-3000 and MultiGauge software. The levels of primer extension (pr.ext.) and primer elongation past the damaged or corresponding undamaged sites (undamaged T, abasic site, or hydantoin) are expressed as a percentage of the total primer termini.

Ancient DNA

Ancient DNA (aDNA) was extracted as described²⁹. Essentially, bone or teeth were ground with mortar and pestle. 10 ml extraction buffer containing 0.45M EDTA (pH 8), 0.5% N-Lauroylsarcosine, 1% Polyvinylpyrrolidone, 50mM DTT, 2.5mM PTB, and 0.25mg/ml Proteinase-K were added to 200mg - 1g of bone powder and incubated for 16 hr at 37 °C under rotation. The remaining bone powder was collected by centrifugation and only the supernatant was used for further processing. aDNA was purified by binding to silica. 40 ml of L2 buffer (5.5 M Guanidinium-isothiocyanate, 25mM NaCl, 100mM Tris [pH 8]) and 50µl of silica suspension were added to 10 ml supernatant and incubated for approximately 30min. The pellet was collected by brief centrifugation, the supernatant discarded, the silica pellet was washed in buffer L2 and once with NewWash (Bio 101, La Jolla, CA). After drying the pellet, the DNA was eluted at 56°C in aliquots of 100 µl TE (10mMTris pH 7.4, 1mM EDTA) Mock extractions were performed alongside all extractions. The final volume of the extract was 100µl. aDNA was amplified by 2-step PCR. Briefly, 2 µl of ancient sample were added to a 20 µl PCR in SuperTaq buffer (HT Biotech) with 1µM primers, 2µM dNTPs as well as 0.5U of SuperTaq or blend and amplified for 28 cycles using primers 34,35 (PCR1). This PCR was set up in a clean room following precautions appropriate for aDNA. PCR1 was then diluted 1 / 20 in a secondary clean room and reamplified for 32 cycles using in-nested primers 36,37 and standard PCR parameters. Blend was prepared by mixing polymerases (activity normalized on undamaged templates) at the following ratios: 90% SuperTaq / 10% mutant polymerases (equivalent amounts of 3A10, 3B5, 3B6, 3B8, 3B10, 3C12, 3D1). Amplifications using SuperTaq alone were compared with amplifications using blend in PCR1 and SuperTaq in PCR2. No template controls were always included to detect contamination.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Md'A was supported by a MRC studentship and a Junior Research Fellowship from Trinity College, Cambridge. A.V and R.W were supported by funds from the NICHD/NIH Intramural Research Program."

References

1. Pääbo S, et al. Genetic Analyses From Ancient DNA. *Annu Rev Genet.* 2004
2. Crameri A, Raillard SA, Bermudez E, Stemmer WP. DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature.* 1998; 391:288–291. [PubMed: 9440693]
3. Ghadessy FJ, Ong JL, Holliger P. Directed evolution of polymerase function by compartmentalized self-replication. *Proc. Natl. Acad. Sci. USA.* 2001; 98:4552–4557. [PubMed: 11274352]
4. Ghadessy FJ, et al. Generic expansion of the substrate spectrum of a DNA polymerase by directed evolution. *Nat Biotechnol.* 2004; 22:755–759. [PubMed: 15156154]
5. Leonard JA, et al. Ancient DNA evidence for Old World origin of New World dogs. *Science.* 2002; 298:1613–1616. [PubMed: 12446908]
6. Jaenicke-Despres V, et al. Early allelic selection in maize as revealed by ancient DNA. *Science.* 2003; 302:1206–1208. [PubMed: 14615538]
7. Anderung C, et al. Prehistoric contacts over the Straits of Gibraltar indicated by genetic analysis of Iberian Bronze Age cattle. *Proc Natl Acad Sci U S A.* 2005; 102:8431–8435. [PubMed: 15941827]
8. Green RE, et al. Analysis of one million base pairs of Neanderthal DNA. *Nature.* 2006; 444:330–336. [PubMed: 17108958]
9. Krings M, et al. Neandertal DNA sequences and the origin of modern humans. *Cell.* 1997; 90:19–30. [PubMed: 9230299]
10. Geigl EM. Inadequate use of molecular hybridization to analyze DNA in Neanderthal fossils. *Am J Hum Genet.* 2001; 68:287–291. [PubMed: 11115383]
11. Plosky BS, Woodgate R. Switching from high-fidelity replicases to low-fidelity lesion-bypass polymerases. *Curr Opin Genet Dev.* 2004; 14:113–119. [PubMed: 15196456]
12. Poinar HN, et al. Metagenomics to Paleogenomics: Large-Scale Sequencing of Mammoth DNA. *Science.* 2005
13. Huang M-M, Arnheim N, Goodman MF. Extension of base mispairs by Taq polymerase: implications for single nucleotide discrimination in PCR. *Nucleic Acids Res.* 1992; 20:4567–4573. [PubMed: 1408758]
14. Miller H, Grollman AP. Kinetics of DNA polymerase I (Klenow fragment exo-) activity on damaged DNA templates: effect of proximal and distal template damage on DNA synthesis. *Biochemistry.* 1997; 36:15336–15342. [PubMed: 9398262]
15. Suzuki M, Baskin D, Hood L, Loeb LA. Random mutagenesis of *Thermus aquaticus* DNA polymerase I: concordance of immutable sites in vivo with the crystal structure. *Proc. Natl. Acad. Sci. USA.* 1996; 93:9670–9675. [PubMed: 8790389]
16. Xia G, et al. Directed evolution of novel polymerase activities: mutation of a DNA polymerase into an efficient RNA polymerase. *Proc. Natl. Acad. Sci. USA.* 2002; 99:6597–6602. [PubMed: 12011423]
17. Ong JL, Loakes D, Jaroslawski S, Too K, Holliger P. Directed evolution of DNA polymerase, RNA polymerase and reverse transcriptase activity in a single polypeptide. *J Mol Biol.* 2006; 361:537–550. [PubMed: 16859707]
18. Matsuda T, et al. Error rate and specificity of human and murine DNA polymerase ϵ . *J Mol Biol.* 2001; 312:335–346. [PubMed: 11554790]
19. Frank EG, et al. Altered nucleotide misinsertion fidelity associated with poliovirus-dependent replication at the end of a DNA template. *Embo J.* 2001; 20:2914–2922. [PubMed: 11387224]

20. Höss M, Jaruga P, Zastawny TH, Dizdaroglu M, Pääbo S. DNA damage and DNA sequence retrieval from ancient tissues. *Nucleic Acids Res.* 1996; 24:1304–1307. [PubMed: 8614634]
21. Lindahl T, Karlstrom. O Heat-Induced Depyrimidination of Deoxyribonucleic Acid in Neutral Solution. *Biochemistry.* 1973; 12:5151–5154. [PubMed: 4600811]
22. Lindahl T, Nyberg B. Rate of Depurination of Native Deoxyribonucleic Acid. *Biochemistry.* 1972; 11:3610. [PubMed: 4626532]
23. Hansen AJ, et al. Crosslinks rather than strand breaks determine access to ancient DNA sequences from frozen sediments. *Genetics.* 2006; 173:1175–1179. [PubMed: 16582426]
24. Bloom JD, Labthavikul ST, Otey CR, Arnold FH. Protein stability promotes evolvability. *Proc Natl Acad Sci U S A.* 2006; 103:5869–5874. [PubMed: 16581913]
25. Zhao H, Giver L, Shao Z, Affholter JA, Arnold FH. Molecular evolution by staggered extension process (StEP) in vitro recombination [see comments]. *Nature Biotechnol.* 1998; 16:258–261. [PubMed: 9528005]
26. Ghadessy FJ, Holliger P. Compartmentalized self-replication: a novel method for the directed evolution of polymerases and other enzymes. *Methods Mol Biol.* 2007; 352:237–248. [PubMed: 17041269]
27. Debbie P, et al. Allele identification using immobilized mismatch binding protein: Detection and identification of antibiotic resistant bacteria and determination of sheep susceptibility to scrapie. *Nucleic Acids Res.* 1997; 25:4825–4829. [PubMed: 9365263]
28. McDonald JP, et al. Novel thermostable Y-family polymerases: applications for the PCR amplification of damaged or ancient DNAs. *Nucleic Acids Res.* 2006; 34:1102–1111. [PubMed: 16488882]
29. Hofreiter M, et al. Evidence for reproductive isolation between cave bear populations. *Curr Biol.* 2004; 14:40–43. [PubMed: 14711412]
30. Tawfik DS, Griffiths AD. Man-made cell-like compartments for molecular evolution. *Nature Biotechnol.* 1998; 16:652–656. [PubMed: 9661199]

Figure 1. Selection scheme

(a) Principle of CSR selection. CSR is based on a simple feedback loop, in which a polymerase replicates only its own encoding gene. Compartmentalization in the aqueous compartments of a water-in-oil emulsion³⁰ serves to isolate individual self-replication reactions from each other. In such a system adaptive gains directly (and proportionally) translate into genetic amplification of the encoding gene. Two independent aqueous emulsion compartments are shown. Polymerases (such as Pol1 (left compartment)) that are capable of utilizing quadruple mismatch primers (AGGG•AGGG, GGTG•GGTG) to replicate only their one encoding gene (*pol1*) produce “offspring” i.e. increase their copy number in the post-selection population, while polymerases like Pol2 (right compartment) that are unable to utilize quadruple mismatch primers disappear from the gene pool (a more detailed step-by-step selection scheme is shown in Supplementary information Figure 1). **(b)** Polymerase activity in PCR of the selected polymerase 3D1 using matched (+), single mismatch (1M: G•A, C•C), double mismatch (2M: CC•CC, CC•CC), quadruple mismatch (4M: AGGG•AGGG, GGTG•GGTG) and abasic site primers (AP). Under the same conditions, wild-type polymerases *Taq*, *Tth*, *Tfi* only yielded amplification products using matched (+) primers. There is substantial primer-dimer formation with the quadruple mismatch primers. The correct size quadruple mismatch amplification product is marked by an arrow. m: ϕ ×174 HaeIII digest marker

Figure 2. Quadruple mismatch extension and template slippage

(a) Enzyme concentrations were normalized on an undamaged template and polymerase activity was assayed for its ability to extend radiolabeled quadruple mismatch primers (M1: GGTG•GGTG, M2: AGGG•AGGG, M3: TTTT•TTTT) and compared to extension of matched primer (M0). Only selected polymerases 3A10 and 3D1 yield extension products with M1-3 but extension products are predominately shorter than M0. Possible primer template configurations and expected main product lengths (N+1) are illustrated in (b). Matched primer-template sequences (M0) at primer 3' end are shown in blue, mismatched and misaligned structures are shown in red. (c) Quadruple mismatch extension of M3 by 3A10, when extended with dTTP only (lane 2, M3_{TT}). Possible primer template configuration for M3 for quadruple mismatch extension is shown with the two incorporated dTs in green (M3_{TT}). When initial dTTP extension is followed by addition of the remaining dNTPs (lane 3), 3A10 yields a significant proportion of full-length extension product as evident by the appearance of bands at (+43), (+44), while addition of dNTPs from the start (lane 4, M3_N) favours the misaligned structure yielding a shorter extension product centered around (+42) as in (A).

Figure 3. Damage tolerance and ancient DNA amplification

(a) Polymerase activity was assayed for its ability to extend a radiolabelled primer. Enzyme concentrations were normalized on an undamaged template and damage bypass activity of *Taq*, 3A10, 3D1 was assayed on templates containing an abasic site (AP), 5-hydroxy-hydantoin (5H) or 5-methyl-5-hydroxy-hydantoin (5M) with polymerase units (U) in the range 0.0001-0.1U (AP, 5H) or 0.002-2U (5M). (b) Percentage of successful PCRs is plotted versus sample dilution for ancient cave bear DNA (GS3-7, 47kyrs old) (left panel) and “modern” DNA (right panel). The blend yields amplicons at sample dilutions where *Taq* no longer yields any despite an apparently substantially reduced affinity for (undamaged) template DNA.

Table 1**Ancient DNA amplification**

Relative PCR success in the presence of limiting amounts of two different ancient cave bear DNA samples (GS3-7 and 366 (60kyrs old)). In eight out nine experiments, the blend of mismatch polymerases yielded more amplicons than *Taq*.

Exp	Sample ^a	Dilution	<i>Taq</i> PCR ^b	Blend PCR ^b	Improvement ^c
1	GS3-7	1/500	24 / 36	28 / 36	+16%
2	GS3-7	1/2000	2 / 24	5 / 24	+150%
3	GS3-7	1/1000	21 / 48	24 / 48	+14%
4	366	1/5	2 / 4	4 / 4	+100%
5	366	1/10	12 / 24	16 / 24	+33%
6	GS3-7	1/1000	12 / 48	14 / 48	+16%
7	GS3-7	1/1000	10 / 24	7 / 24	-30%
8	GS3-7	1/1000	12 / 48	13 / 48	+8%
9	GS3-7	1/1000	12 / 48	13 / 48	+8%
Total			107 / 304	124 / 304	15%

^a 366 derives from the Herdengel cave (Austria) (ca. 60 kyrs old); GS3-7 derives from the Gamsulzen cave (Austria) (ca. 47 kyrs old)

^b successful PCR amplifications / total PCR amplifications per experiment

^c improvement Blend vs *Taq* (%)